Codes and infinite words*

J. Devolder [†]   M. Latteux[†]   I. Litovsky[‡]   L. Staiger[§]

Abstract

Codes can be characterized by their way of acting on infinite words. Three kinds of characterizations are obtained. The first characterization is related to the uniqueness of the factorization of particular periodic words. The second characterization concerns the rational form of the factorizations of rational words. The third characteristic fact is the finiteness of the number of factorizations of the rational infinite words. A classification of codes based on the number of factorizations for different kinds of infinite words is set up. The obtained classes are compared with the class of ω-codes, the class of weakly prefix codes and the class of codes with finite deciphering delay. Complementary results are obtained in the rational case, for example a necessary and sufficient condition for a rational ω-code to have a bounded deciphering delay is given.

Résumé: La factorisation des mots infinis permet de caractériser les codes parmi les langages de mots finis. Les critères obtenus sont de trois types. Le premier critère est relatif à l'unicité de la factorisation de certains mots périodiques. Le second concerne la forme des factorisations des mots rationnels. Finalement, seuls les codes nous assurent de la finitude du nombre de factorisations des mots rationnels. Les codes sont classifiés selon le nombre de factorisations de certains types de mots infinis. Les classes obtenues sont étudiées et comparées avec les classes déjà définies de ω-codes, de codes faiblement préfixes et de codes à délai borné. Des résultats complémentaires sont obtenus dans le cas rationnel, en particulier il est donné une condition nécessaire et suffisante pour qu'un ω-code rationnel soit à délai borné.

Introduction

Codes, which are defined as the bases of free submonoids of monoids of (finite) words [1] were initially introduced by Schützenberger [19] in 1955. Since then, the

study of some classes of codes, specially from the point of view of an easy decoding, has been very active. Here we study codes, and classes of codes from the particular point of wiew of decoding infinite words. In this respect, the interesting codes are those for which every infinite word has at most one factorisation. We shall refer to these codes as ω-codes. It was shown by Levenshtejn [12] that, for a finite code, any infinite word has at most one factorisation iff this code has a bounded deciphering delay. For infinite codes the situation is more complicated. It turns out that the class of ω-codes (initially called ifl-codes by Staiger) properly contains the class of codes having a finite deciphering delay, which in one's turn properly contains the class of codes having a bounded deciphering delay [20]. The most interesting codes are codes with bounded deciphering delay, because they allow an easy decoding of finite and infinite words. We give at the end of this paper an interesting necessary and sufficient condition for a rational ω-code to have a bounded deciphering delay.

Although arbitrary codes may give several factorizations of infinite words, codes can be characterized by their way of acting on infinite words. This is the purpose of the first section. Indeed, a language C is a code if and only if, for every word v of C^+, the periodic infinite word v^ω has a single factorization over C. Codes are also characterized by the form of the factorizations of ultimately periodic words, and also by the fact that the number of factorizations of an arbitrary ultimately periodic word is finite. As an application, it is shown that the usual notion of code with bounded deciphering delay coincide with the notion defined in [20].

So, codes and ω-codes are characterized in terms of infinite words. It is obvious that a language C is a code if no infinite word has uncountably many factorizations over C. Having this fact in mind, we set up a classification of codes based on the number of factorisations for different kinds of infinite words. If C denotes a code, the kinds of infinite words that we consider are the following ones: periodic words of the form u^ω with $u \in C^+$, periodic words, ultimately periodic words and any infinite words. This leads to consider the class C of codes, the class Π of π-codes, the class W of weakly prefix codes, the class I of ω-codes (I as "iflcode").

These classes are compared with each other, and also compared with the class B of codes having a bounded deciphering delay, the class D of codes having a finite deciphering delay, the class V of circular codes (V as "very pure"), the class S of suffix codes. The results can be summarized by the following strict inclusions $\mathbb{B} \subset \mathbb{D} \subset \mathbb{I} \subset \mathbb{W} \subset \Pi \subset \mathbb{C}$, $\mathbb{V} \subset \mathbb{W}$, $\mathbb{S} \subset \Pi$, and by the next array which indicates the maximal number of factorizations according to the type of infinite words and the class of codes, when the alphabet is countable and has at least two elements. In this array, the stars $\diamond$ point out the characteristic properties, and ∞ denotes $\text{Card}(\mathbb{R})$: a noncountable infinity of factorizations is possible.

words languages	u^ω ($u \in C^+$)	u^ω	uv^ω	any
ω-codes	1	1	1	1$\diamond$
weakly prefix	1	1	1$\diamond$	∞
π-codes	1	1$\diamond$	finite	∞
codes	1$\diamond$	finite $\diamond$	finite $\diamond$	∞

In the second section, we give characterizations for the classes $\mathbb{W}$, Π and S and we prove the announced inclusions. Using the inclusions $\mathbb{V} \subset \mathbb{W}$, $\mathbb{S} \subset \Pi$ and the composition of codes, one can easily construct ω-codes, weakly prefix codes and π-codes. The second section terminates by some examples which enable us to fulfill the array.

In the last section, we examine the modifications holding when C is a rational language. Every infinite word has then a finite bounded number of C-factorizations whenever C is a code. The notion of ω-code coincides with the notion of weakly prefix code in the rational case. We give also a new interesting necessary and sufficient condition for a rational ω-code to have a finite deciphering delay. This condition $C^\omega \cap C^* \operatorname{Adh}(C) = \emptyset$ can be easily checked. As expected, it is decidable whether a rational language belongs to any class $\mathbb{B}, \mathbb{I}$ (i.e. $\mathbb{W}$) or Π.

Notations and basic definitions:

In the following, we consider an alphabet (finite or not) A, the set A^* (resp. A^ω) of all finite (resp. infinite) words over A, the set A^+ which denotes the language $A^* - \varepsilon$, where ε is the empty word. The length of a word u is denoted by $|u|$. The symbol $\leq$ (resp $<$) denotes the relation between words "is a (resp. strict) prefix of". The left quotient of a word u by a word v is denoted by $v^{-1}u$.

Two words x and x' are said to be *conjugate* if there exist u and v such that $x = uv$ and $x' = vu$. A word $z \in A^+$ is *primitive* if $z = u^n$ implies $n = 1$.

Given a language $C \subset A^+$, the submonoid generated by C is the language $C^* = \{v_1 \ldots v_n | n \geq 0, v_i \in C, 1 \leq i \leq n\}$ and C^ω stands for the set of infinite words obtained by concatenation of an infinite sequence of words of C : $C^\omega = \{v_0 v_1 v_2 \ldots | v_i \in C, i \geq 0\}$. A *C-factorization* of a word $v \in C^*$ is a sequence of words of C : $(v_1, \ldots, v_n)$ such that $v = v_1 \ldots v_n$. A *C-factorization* of a word $v \in C^\omega$ is a sequence of words of C : $(v_0, v_1, v_2, \ldots)$ such that $v = v_0 v_1 v_2 \ldots$.

An infinite word w is said to be *ultimately periodic* if there exist finite words u and v such that $w = uv^\omega$. It is said to be *periodic* if u can be chosen equal to ε.

Given a language $C \subset A^+$ we shall often consider a bijection φ between an alphabet X and the language C. This mapping can be extended to X^* as a morphism $\varphi : X^* \to C^*$. This morphism is said to be a *coding morphism for C* (even if it is not injective). The mapping φ can also be extended to $X^\omega (\varphi(z_0 z_1 \ldots)$ is the word $\varphi(z_0)\varphi(z_1) \ldots)$. These extension agree with the composition of functions of words of X^* (resp. X^ω) and the set of C-factorizations of words of C^* (resp C^ω). Thus a C-factorization of $u \in C^*$ (resp : $u \in C^\omega$) will be represented by an element of X^* (resp: X^ω).

Definitions: Let C be a language $\subset A^+$.
  - C is a code if and only if $\forall u, v \in C \quad uC^* \cap vC^* \neq \emptyset \Rightarrow u = v$
  - C is a prefix code if and only if $\forall u, v \in C \quad u \leq v \Rightarrow u = v$
  - C is an ω-code if and only if $\forall u, v \in C \quad uC^\omega \cap vC^\omega \neq \emptyset \Rightarrow u = v$ [20].

These definitions can be expressed in terms of morphisms. Let φ be any coding morphism for C.
  - C is a code if and only if $\varphi : X^* \to C^*$ is injective.
  - C is an ω-code if and only if $\varphi : X^\omega \to C^\omega$ is injective.

Recall that ω-codes are codes and that prefix codes are ω-codes. Using coding morphisms, it is easily seen that a composition theorem holds for codes [1] and ω-codes. Namely, let C be a language $\subset X^+$ and $\varphi : X^* \to A^*$ be a coding morphism for a language $D = \varphi(X) \subset A^+$, if C and D are codes (resp: ω-codes), $\varphi(C)$ is a code (resp: ω-code).

1 Characterizations of codes

In this section, three kinds of characterizations for codes are obtained: the first kind concerns the words which have only one C-factorization, the second is related to the form of the C-factorizations of ultimately periodic words, the last give a bound for the number of C-factorizations of a given ultimately periodic word.

We define now some notations and give some lemmata used in the proof of the main theorem. Let $\varphi : X^* \to C^*$ be a coding morphism for C.

Lemma 1.1 *If $C \subset A^+$ is a code, for every word $v \in A^+$, there exists at most one primitive word $z \in X^+$ such that $\varphi(z) \in v^+$.*

Proof. If $\varphi(z) = v^n$ and $\varphi(z') = v^m, \varphi(z^m) = \varphi(z'^n)$. Thus $z^m = z'^n$ (φ injective) and then $m = n$ and $z = z'$ if z and z' are primitive words. □

Lemma 1.2 *If $yz^\omega \in X^\omega$ is a C-factorization of uv^ω (where v is assumed to be a primitive word), $\varphi(z)$ is a power of a conjugate of v.*

Lemma 1.3 *Let us consider $x \in X^\omega$ such that $\varphi(x)$ is ultimately periodic. There exist $y, z \in X^*, t \in X^\omega$ such that $x = yzt$, and $\varphi(x) = \varphi(y)\varphi(z)^\omega$.*

Proof. Let x be the C-factorization: $u_1, u_2, \ldots, u_p, \ldots$ of an ultimately periodic word uv^ω. Since v is of finite length there exist i, j, k, m such that $k < m$, $u_1 \ldots u_k = uv^i w$ and $u_1 \ldots u_m = uv^{i+j}w$ where w is a prefix of v. The word $v' = w^{-1}v^j w$ belongs to C^+ and $uv^\omega = u_1 \ldots u_k v'^\omega$. □

Lemma 1.4 *If $C \subset A^+$ is a code, for every word $v \in C^+$, the word v^ω has only one C-factorization.*

Proof. Let us consider $v \in C^+ : v = v_1 v_2 \ldots v_n$ with $v_i \in C$ such that v^ω has two distinct C-factorizations: $v^\omega = (v_1 v_2 \ldots v_n)^\omega = u_1 u_2 \ldots u_p \ldots$ (where $\forall i\ u_i \in C$). Without loss of generality we may assume that $v_1 \neq u_1$. As in the proof of lemma 1.3, there exist i, j, k, m such that $k < m$, $u_1 \ldots u_k = v^i w$ and $u_1 \ldots u_m = v^{i+j}w$ where w is a prefix of v. Then the word $v^{i+j}w = u_1 \ldots u_m = (v_1 \ldots v_n)^j u_1 \ldots u_k$ has two distinct C-factorizations. C is not a code. □

Lemma 1.5 *Consider $C \subset A^+$ such that every word of the form w^ω with $w \in C^+$ has exactly one C-factorization. For all words $u, v \in A^+$, every C-factorization of the word uv^ω is ultimately periodic.*

Proof. Let us consider a C-factorization x of the word $uv^\omega \in C^\omega$. From lemma 1.3, there exist $y, z \in X^*, t \in X^\omega, v' \in C^+$ such that $x = yzt, \varphi(z) = v', \varphi(t) = v'^\omega$. By hypothesis, the word v'^ω has a single C-factorization. Since $\varphi(z^\omega) = v'^\omega = \varphi(t)$, we have $t = z^\omega$ and then x is ultimately periodic. □

Lemma 1.6 *Let C be a code $\subset A^+$. Consider words u and v of A^+. The set of C-factorizations of uv^ω is finite.*

Proof. Let us consider $uv^\omega \in C^\omega$. Assume that v is primitive. Denote by $V = \{v_i | i \in I\}$ the set of conjugates v_i of v such that $v_i^+ \cap C^+ \neq \emptyset$. Since C is a code, we can denote by z_i the primitive word such that $\varphi(z_i) \in v_i^+$ and n_i the corresponding power of $v_i : \varphi(z_i) = v_i^{n_i}$. We consider the equivalence relation on V:

$$v_i \simeq v_j \Leftrightarrow z_i \text{ and } z_j \text{ are conjugate.}$$

Since $\varphi(z_i)$ and $\varphi(z_j)$ are conjugate, it is clear that $n_i = n_j$ whenever $v_i \simeq v_j$.

Let F be the set of C-factorizations of uv^ω. We shall prove that $\mathrm{Card}(F) \leq \Sigma n_i$, where only one n_i by $\simeq$ class is taken.

Since C is a code, from lemma 1.5, every C-factorization of uv^ω is ultimately periodic, hence of the form yz^ω with z primitive; from lemma 1.2, there exists a conjugate v_i of v such that $\varphi(z) \in v_i^+$. Then the set F of C-factorizations of uv^ω satisfies $F = \cup(F \cap X^* z_i^\omega)$. Since $X^* z_i^\omega = X^* z_j^\omega$ when z_i and z_j are conjugate, the previous union has only N terms, where N denotes the number of classes of $\simeq$.

It remains to prove that $\mathrm{Card}(F \cap X^* z_i^\omega) \leq n_i$. Consider $y' z_i^\omega, y'' z_i^\omega$ and $yz_i^\omega \in F$, such that $|\varphi(y)| = \inf\{|\varphi(u)| | uz_i^\omega \in F\}$. Since $\varphi(y)\varphi(z_i)^\omega = \varphi(y')\varphi(z_i)^\omega, \varphi(z_i) \in v_i^+$ and v_i is primitive, one has $\varphi(y') = \varphi(y)v_i^{h'}$ for some h'. One has also $\varphi(y'') = \varphi(y)v_i^{h''}$ for some h''. If $h' = kn_i + h''$, $\varphi(y'' z_i^k) = \varphi(y)v_i^{h''+kn_i} = \varphi(y)v_i^{h'} = \varphi(y')$. Since C is a code $y'' z_i^k = y'$ and then $y'' z_i^\omega = y' z_i^\omega$. The number of elements of $F \cap X^* z_i^\omega$ is then at most the number n_i of integers modulo n_i. □

The following theorems give the characterization of codes. For convenience, theorem 1.7 gives the characterizations related to periodic words ant theorem 1.8 gives those related to ultimately periodic words.

Theorem 1.7 *Let C be a language $\subset A^+$. The following assertions are equivalent:*

1. *C is a code,*

2. *for every $u \in C^+$, u^ω has a single C-factorization,*

3. *every C-factorization of each periodic infinite word is ultimately periodic,*

4. *each periodic infinite word has a finite number of C-factorizations.*

Theorem 1.8 *Let C be a language $\subset A^+$. The following assertions are equivalent:*

1 C is a code,

3' every C-factorization of each ultimately periodic infinite word is ultimately periodic,

4' each ultimately periodic infinite word has a finite number of C-factorizations.

Proof. $1 \Rightarrow 2$: lemma 1.4; $2 \Rightarrow 3'$: lemma 1.5; $1 \Rightarrow 4'$: lemma 1.6; $3' \Rightarrow 3, 4' \Rightarrow 4$: clear; $3 \Rightarrow 1$ and $4 \Rightarrow 1$: If C is not a code, there exists a word u which has two distinct C-factorizations. There exist y and $z \in X^+, y \neq z$, such that $\varphi(y) = \varphi(z) = u$. Without loss of generality, one can assume that the first letters of y and z are different, then a bijection Ψ between $\{0, 1\}$ and $\{y, z\}$ gives a bijective morphism $\Psi : \{0, 1\}^\omega \to \{y, z\}^\omega$ and the elements of $\{y, z\}^\omega$ are distinct C-factorizations of u^ω. The word u belongs to C^+ and u^ω has a non-countable set of C-factorizations; hence also a non-countable number of non-ultimately periodic C-factorizations. □

Remarks:
 – From lemma 1.3, in the property 4', one can replace: "each ultimately periodic infinite word" by "each ultimately periodic infinite word of the form uv^ω with $u, v \in C^+$".

– A periodic infinite word can have a nonperiodic C-factorization even if C is a (prefix) code. For example: if $C = \{a, ba\}$, the C-factorization of $(ab)^\omega$ is not periodic.

Property 3' of codes has been used to give characteristic properties of precircular codes [7]. The characterizations 3 and 3' can be used to prove composition theorems for weakly prefix codes and for π-codes. As an application of property 2, it can be easily seen that a code C is always minimal in the family of ω-generators of C^ω (i.e. languages R such that $R^\omega = C^\omega$). We give here another application of property 2.

Application:

In [20] the following notion of delay of decipherability was introduced: a language $C \subset A^+$ is said to have a *finite delay of decipherability* if

$$\forall v \in C \; \exists m(v) \geq 0 \quad vC^{m(v)}A^\omega \cap C^\omega \subset vC^\omega.$$

Remark: A language with a finite delay of decipherability in this sense is not necessarily a code, as it can be seen for $C = \{a, a^2\}$. The language $C = \{a^2, a^3, b\}$ is another more complicated example (it is not a code but $m(b) = 0$ and $m(a^2) = m(a^3) = 1$).

Some authors use another notion of finite deciphering delay [1], [5], which is in fact a notion of bounded deciphering delay [10]. Here, we say that:
– a language $C \subset A^+$ is said to have a *finite deciphering delay* if

$$\forall v \in C \quad \exists m(v) \geq 0 \quad \forall v' \in C \; (vC^{m(v)}A^\omega \cap v'C^\omega \neq \emptyset \Rightarrow v = v')$$

or equivalently if

$$\forall v \in C \quad \exists m(v) \geq 0 \quad \forall v' \in C \; (vC^{m(v)}A^* \cap v'C^* \neq \emptyset \Rightarrow v = v')$$

A language which has a finite deciphering delay is a code [1] and clearly has a finite delay of decipherability in the sense of [20]. Thus the notion of finite deciphering delay is stronger than the notion defined by Staiger. We shall see that these notions coincide for codes.

Proposition 1.9 *Every code which has a finite delay of decipherability is an ω-code.*

Proof. Consider $v, v' \in C$ such that $vC^\omega \cap v'C^\omega \neq \emptyset$. For $n \geq \max(m(v), m(v'))$ and $w \in vC^\omega \cap v'C^\omega$, there exist $u, u' \in C^n$ such that vu and $v'u'$ are prefixes of w. If vu is a prefix of $v'u'$, $(v'u')^\omega \in vC^{m(v)}A^\omega \cap C^\omega$, thus $(v'u')^\omega \in vC^\omega$. Since $v'u' \in C^+$, from characterization 2, $v = v'$. Hence C is an ω-code.   □

Proposition 1.10 *Every code which has a finite delay of decipherability has a finite deciphering delay.*

Proof. Let v and $v' \in C$ and assume that $vC^{m(v)}A^* \cap v'C^*$ is not empty. Consider $w \in vC^{m(v)}A^* \cap v'C^*$. The word wv^ω belongs to $vC^{m(v)}A^\omega \cap C^\omega$ and then belongs to vC^ω and to $v'C^\omega$, from the previous proposition we obtain that $v = v'$.   □

Remarks:
– In a same way, a code satisfying: $\exists m \geq 0 \quad \forall v \in C \quad vC^{m(v)}A^{\omega} \cap C^{\omega} \subseteq vC^{\omega}$
is a code with a bounded deciphering delay, that is to say:

$$\exists m \geq 0 \forall v \in C \forall v' \in C(vC^m A^* \cap v'C^* \neq \emptyset \Rightarrow v = v').$$

– The two notions of finite and bounded delay do not coincide in general, although they are equivalent in the regular case [20].
– The notions of ω-code and code with a (finite or bounded) deciphering delay coincide in the case of finite codes [12] [5]; these classes do not coincide when regular codes are considered [20] . We give in section 3 a necessary and sufficient condition for a rational ω-code to have a finite deciphering delay.

2   Study of some special codes - examples.

Weakly prefix codes were defined by Capocelli [5]:

Definition: A code $C \subset A^+$ is a weakly prefix code if and only if

$$\forall u, v, w \in A^* \quad (w, wu, uv, vu \in C^* \Rightarrow u \in C^*).$$

Notice that this definition is equivalent to the next:
A language $C \subseteq A^+$ is a *weakly prefix code* if and only if C is the base of a monoid M satisfying the condition:

$$\forall u, v, w \in A^* \quad (w, wu, uv, vu \in M \Rightarrow u \in M).$$

Proof. It is sufficient to prove that a monoid M which satisfies the required condition is stable [1]. If the words w, wu, uv', v' belong to M, the words $w, wu, uv'w, v'wu$ belong also to M. Let $v = v'w$. The words w, wu, uv, vu belong to M and then u belong to M, M is stable. $\square$
Clearly, prefix codes are weakly prefix codes.
Let us recall some definitions. A language $C \subset A^+$ is a circular code [11] [1] if and only if

$$\forall n, p \geq 0 \forall u_0, \ldots, u_{n-1}, v_0, \ldots, v_{p-1} \in C \forall t \in A^* \forall s \in A^+ \text{ such that } v_0 = ts$$

$$(u_0 \ldots u_{n-1} = sv_1 \ldots v_{p-1}t \Rightarrow n = p \ t = \varepsilon \text{ and } \forall i \ u_i = v_i).$$

A monoid $M \subset A^*$ is a very pure monoid if and only if

$$\forall u, v \in A^* \quad (uv, vu \in M \Rightarrow u, v \in M).$$

It is known that a language C is a circular code if and only if C is the base of a very pure monoid [16].
Clearly, the class $\mathbf{V}$ of circular codes is an interesting subclass of the class $\mathbf{W}$ of weakly prefix codes. But the inclusion $\mathbf{V} \subset \mathbf{W}$ is strict: for example, $\{ab, ba\}$ is a (weakly) prefix code but is not a circular code.
The next proposition characterizes weakly prefix codes in terms of infinite words.

Theorem 2.1 *Let C be a language $\subset A^+$. The following assertions are equivalent:*

1. *C is a weakly prefix code*

2. *for every $u, v \in C^+$ uv^ω has a single C-factorization*

3. *each ultimately periodic infinite word has at most one C-factorization.*

Proof. $1 \Rightarrow 2$: Notice that, since C is a code, any C-factorization of an ultimately periodic word is ultimately periodic. Assume now that $uv^\omega = u'v'^\omega$ and $u, u', v, v' \in C^+, |u| \le |u'|$. If the two C-factorizations are distinct, we can assume that $u = u_1 \dots u_n$ and $u' = u_1' \dots u_p'$ with $u_1 \ne u_1'$. If $u' = uw$ where $w \in C^*, C$ is not a code. Then suppose that $u' = uw$ where $w \notin C^*$. Taking appropriate powers of v and v', we can assume that $|v| = |v'| > |w|$. Then $v = ww'$ and $v' = w'w$ for some word w'. We have $u, uw, ww', w'w \in C^*$ but $w \notin C^*$, a contradiction with C weakly prefix.

$2 \Rightarrow 1$: If C is not weakly prefix, there exist u, v, w such that $u \notin C^*, w, wu, uv, vu \in C^*$. Hence $w(uv)^\omega$ has two distinct C-factorizations.

$3 \Rightarrow 2$: Clear. $2 \Rightarrow 3$: Clear from lemma 1.3.   □

As a consequence we obtain:

Corollary 2.2 *ω-codes are weakly prefix codes.*

The converse is not true in general. Let $C = \{ab\} \cup \{ab^n ab^{n+1} | n \ge 1\}$. This example presents a weakly prefix (circular) code C which is not an ω-code, but such that every proper subset of C is an ω-code. This shows a difference between ω-codes and weakly prefix codes since a language C is clearly a weakly prefix code iff every finite subset of C is a weakly prefix code. This example shows also that $\mathbb{V}$ and $\mathbb{W}$ are not included in the class $\mathbb{I}$ of ω-codes; $\mathbb{I}$ is neither included in $\mathbb{V}$ (consider the prefix code $\{ab, ba\}$).

Now, we study a type of codes which take place between codes and weakly prefix codes. Indeed, such a type of codes exists. From theorem 1.7, if C is a code, for every $u \in C^+$ u^ω has a single C-factorization. But it is not possible to replace "$u \in C^{+}$" by "$u \in A^{+}$". This observation was already made by Karhumäki in connection with theorem 3.3 of [10], however the example given there, $\{ab, aba, baba\}$ is not a code. By contrast, the language $C = \{a, aaba, abaaba\}$ is a code and the word $(aab)^\omega$ has two C=factorizations.

In theorem 2.1, it is not possible to replace "ultimately periodic" by "periodic": a language C may no longer be a weakly prefix code even if every periodic infinite word has at most one C-factorization. For example, let $C = \{ab, aba, ba^2\}$. The word $ab(aba)^\omega = aba(ba^2)^\omega$ is the only word which has at least two C-factorizations beginning by two different words. Thus every periodic word has at most one C-factorization. Note that C is a suffix code.

Thus theorem 1.7 and 2.1 do not study uniqueness of the factorization of periodic words. Then it is natural to try to characterize codes which factorize infinite periodic words in a single manner. For sake of convenience these codes are called π-*codes* here. Note that the three-element codes which are not π-codes have been studied by Karhumäki and called periodic codes [10].

Definition: A language $C \subset A^+$ is said to be a π-code if each periodic infinite word has at most one C-factorization.

Theorem 1.7 ensures that a π-code is a code. We have seen an example showing that the converse is false. As for weakly prefix codes, a technical characterization

of π-codes can be obtained. One can prove that a code $C \subset A^+$ is a π-code if and only if C satisfies the property:

(P) $\forall u, v, w, \beta \in A^*$ such that $wuvu < \beta^\omega$ and $|u| \geq |\beta|$, one has:

$$w, wu, uv, vu \in C^* \Rightarrow u \in C^*.$$

Proof. Let u, v, w, β such that $wuvu < \beta^*, |u| \geq |\beta|$ and $w, wu, uv, vu \in C^*$. We can assume β primitive, then u has a single interpretation over β : there exist a single $i \geq 0$, a single suffix of $\beta : \beta'$, a single prefix of $\beta : \beta''$ such that $u = \beta'\beta^i\beta''$. Then $uv = \beta'\beta^j\beta\beta'^{-1}$ and $vu = \beta''^{-1}\beta\beta^j\beta''$ for some j. Hence $\beta^\omega = w(uv)^\omega = wu(vu)^\omega$. Since C is a π-code, the word β^ω h as at most one C-factorisation therefore $u \in C^*$.

Conversely, let β^ω be a periodic word having two distinct C-factorisations: $(w_1, w_2, \ldots)$ and $(w_1', w_2', \ldots)$. We can assume that $w_1 \neq w_1'$. Denote $\beta^\omega = u_1 u_2 \ldots$ where $u_i = \beta$ for each i.

We can consider (when exists) p_i such that $w_1 \ldots w_{p_i-1} < u_1 \ldots u_{i-1} \leq w_1 \ldots w_{p_i} \leq u_1 \ldots u_i$. There exist a word α and infinitely many i such that $w_1 \ldots w_{p_i} = u_1 \ldots u_{i-1}\alpha$. In the sequel, m and n denote such indices p_i. In a same way, there exist a word α' and infinitely many i such that there exists q_i satisfying $w_1 \ldots w_{q_i-1} < u_1 \ldots u_{i-1} \leq w_1 \ldots w_{q_i} \leq u_1 \ldots u_i$ and $w_1' \ldots w_{q_i}' = u_1 \ldots u_{i-1}\alpha'$. In the sequel, m' and n' denote such indices q_i.

Let us choose m, m', n, n' such that $w_1 \ldots w_m < w_1' \ldots w_{m'}' < w_1 \ldots w_n < w_1' \ldots w_{n'}'$. Let $w = w_1 \ldots w_m, wu = w_1' \ldots w_{m'}', wuy = w_1 \ldots w_n, wuyz = w_1' \ldots w_{n'}'$. The choice of m' can be done such that $|u| \geq |\beta|$. We have: $uy = \beta'^p \in C^+$ and $yz = \beta''^p \in C^+$, where β' and β'' are conjugate with β. Let $v = y(uy)^{q-1}; uv \in C^+$ and $vu = (yz)^p$, then $vu \in C^+$. The words w, wu, uv, vu belong to C^*, therefore u belongs to C^*, which gives a contradiction with "C is a code" since $w_1 \neq w_1'$. □

In this characterisation, the condition "C is a code" cannot be suppressed. For example, let $C = \{ba, b, abc, bc\}$. The monoid C^* is not free and the condition (P) is satisfied.

From theorem 2.1, it is clear that weakly prefix codes are π-codes. Surprisingly, the family of π-codes contains a well-known subfamily: the family S of suffix codes. This fact is obtained as a consequence of the next interesting characterization of suffix codes.

Proposition 2.3 *A language $C \subset A^+$ is a suffix code if and only if every C-factorization of a periodic infinite word is periodic.*

Proof. If C is not a suffix code, there exist $v' \in A^+, u, v \in C$ such that $v = v'u$. The word uv^ω is periodic and has a non periodic C-factorization.

Conversely, consider a suffix code $C, \varphi : X^* \to C^*$ a coding morphism for C and β a primitive word such that $\beta^\omega \in C^\omega$. Consider a C-factorization of β^ω. From lemma 1.2 and theorem 1.7, this factorization can be written yz^ω and there exists a conjugate of $\beta : \beta'$ such that $\varphi(z) = \beta'^n$ for some n. Since $\beta^\omega = \sigma\beta'^\omega, \varphi(y) = \sigma\beta'^k$ for some k. Then $\varphi(y)$ is a suffix of $(\beta'^n)^+$, and since C is a suffix code, y is a suffix of z^+. Hence the considered factorization is periodic. □

In these conditions, $\beta^\omega = v^\omega$ for some v in C^+ and from theorem 1.7, the C-factorisation of β^ω is unique. So we have:

Corollary 2.4 *Suffix codes are π-codes.*

Remarks:

 – The inclusion $S \subset \Pi$ is strict: the π-code $\{a, ba\}$ is a not a suffix code.
 – A code with a finite left deciphering delay (even delay 1) is not always a π-code.

For example: the word $(abc)^\omega$ has two C-factorisations when $C = \{a, ab, cab, bca\}$.

 – S is not included in W : $\{c, ca, aba, ba^2\}$ is a suffix code which is not weakly prefix.

As an application of theorems 1.8 and 2.1, a composition property for weakly prefix codes and π-codes can be obtained:

Proposition 2.5 *Let C be a language $\subset X^+$ and $\varphi : X^\circ \rightarrow A^\circ$ be a coding morphism for a language $D = \varphi(X) \subset A^+$. If C and D are weakly prefix codes, $\varphi(C)$ is a weakly prefix code. If C is a weakly prefix code and D a π-code, $\varphi(C)$ is a π-code.*

Remark: In proposition 2.5, for $\varphi(C)$ to be a π-code, the request property "C weakly prefix code" cannot be replaced by the other one "C π-code". For example, $C = \{c, ca, aba, baa\}$ is a π-code but not a weakly prefix code (the word $c(aba)^\omega$ has two C-factorisations). Let $\varphi(a) = ac, \varphi(b) = b, \varphi(c) = c$. The code $D = \{ac, b, c\}$ is prefix but $\varphi(C) = \{c, cac, acbac, bacac\}$ is not a π-code since the word $(cacba)^\omega$ has two C-factorisations.

In the following, we give some examples of π-codes and weakly prefix codes for which there exists a word w_0 which has infinitely many factorisations. The set of factorisations of w_0 may be countable or not countable. The last example allows us to fulfill the array given in the introduction.

Example 2.1 *Let $C_1 = \{aba^2b^2a^3b^3 \ldots a^nb^na^{n+1}|n \geq 1\}$, $C_2 = \{b^pa^qb^p|0 < p < q\}$ and consider $C = C_1 \cup C_2$. The language C is a suffix code and thus a π-code, but C is not a weakly prefix code since for example, the word $aba^2b^2a^3b^3(a^4b^4)^\omega$ has two C-factorizations*

The word $w_0 = aba^2b^2 \ldots a^nb^na^{n+1}b^{n+1} \ldots$ has a countable infinity of C-factorizations and every word has a countable (finite or not) number of C-factorizations.

Example 2.2 *Let $A = \{a, b\}, C = \{uab^n||u| = n, n \geq 0, |u|_a = 0 \text{ or } 1\}$ ($|u|_a$ denotes the number of occurrences of a in u). Clearly C is a suffix code thus a π-code. Since the word $w = bab.bab.(b^4ab^4.bab)^\omega = bab^2ab^4.(bab.b^3ab^4)^\omega$ has two C-factorizations, C is not a weakly prefix code. We shall see that there exists a word w_0 which has a noncountable infinity of C-factorizations.*

Let w_0 be the word: $ab^{i_0}ab^{i_1} \ldots ab^{i_n} \ldots$ where $i_0 = 0, i_1 = 1, i_{n+2} = i_{n+1} + i_n + 1$ for every $n \geq 0$. Let us prove that, for every factorisation of $w_0 : w_0 = uv$, the word v has at least two C-factorizations. In fact, $v \in x(v)C^\omega \cap y(v)C^\omega$ for two different words: $x(v)$ and $y(v)$ of C. Let $v = b^{j_0}ab^{i_n}ab^{i_{n+1}} \ldots$ with $0 \leq j_0 \leq i_{n-1}$. Then $v = b^{j_0}ab^{j_0}.b^{j_1}ab^{j_1} \ldots . b^{j_h}ab^{j_h} \ldots$ where $j_{h+1} = i_{n+h} - j_h$ and j_h satisfies $0 \leq j_h \leq i_{n+h}$ for every $h \geq 0$; let us set $x(v) = b^{j_0}ab^{j_0}$. The word v has also the other C-factorisation: $v = b^{j_0}ab^{i_n}ab^{k_0}.b^{k_1}ab^{k_1} \ldots . b^{k_h}ab^{k_h} \ldots$ where $k_0 = j_0 + i_n + 1, k_{h+1} = i_{n+1+h}k_h$ and k_h satisfies $0 \leq k_h \leq i_{n+1+h}$ for every $h \geq 0$; let us set $y(v) = b^{j_0}ab^{i_n}ab^{k_0}$. We have: $v \in x(v)C^\omega \cap y(v)C^\omega$.

Then an injective mapping δ form $\{0, 1\}^N$ into the set of C-factorizations of w_0 can be defined next way: let $\beta = (\beta_n)_n \in \{0, 1\}^N$, $\delta(\beta) = (z_n)_n$ where $z_0 = x(w_0)$

if $\beta_0 = 0$ and $z_0 = y(w_0)$ if $\beta_0 = 1, z_n = x((z_0 z_1 \ldots z_{n-1})^{-1} w_0)$ if $\beta_n = 0$ and $z_n = y((z_0 z_1 \ldots z_{n-1})^{-1} w_0)$ if $\beta_n = 1$. So w_0 has a noncountable infinity of C-factorizations. □

Example 2.3 *Let* $A = \{u_i | i \geq 0\}$ *and* $C = C_1 \cup C_2$ *where* $C_1 = \{u_0 u_1 u_2 \ldots u_{2^i} | i \geq 0\}$ *and* $C_2 = \{u_{2^i 3^j + 1} \ldots u_{2^i 3^{j+1}} | i, j > 0\}$. *Since the mapping:* $(i, j) \mapsto 2^i 3^j$ *is injective, it can be shown that* C *is a weakly prefix code. Every word has a countable (finite or not) number of* C-*factorizations and there exists a word which has a countable infinity of* C-*factorizations. Indeed the word* $w_0 = u_0 u_1 u_2 \ldots u_n \ldots$ *has a countable infinity of* C-*factorizations since the* C-*factorizations of* w_0 *are of the form:* $(u_0 \ldots u_{2^i})(u_{2^i+1} \ldots u_{2^i 3})(u_{2^i 3+1} \ldots u_{2^i 3^2}) \ldots (u_{2^i 3^j+1} \ldots u_{2^i 3^{j+1}}) \ldots$ *for some* $i \geq 0$. □

Example 2.4 *Let* $A = \{u_i | \geq 1\}$, $C = \{u_n \ldots u_{2n-1} | n \geq 1\} \cup \{u_n \ldots u_{2n} | n \geq 1\}$. *We show that* C *is a weakly prefix code such that there exists a word which has a noncountable infinity of* C-*factorizations.*

Let $w_0 = u_1 u_2 \ldots u_n \ldots$. As in example 2.2, it can be easily verified that w_0 has a noncountable infinity of C-factorizations. Let w be a word which has two C-factorizations δ and δ' beginning by two different words. Then δ and δ' begin by $u_n \ldots u_{2n-1}$ and $u_n \ldots u_{2n}$ for some n. The second words of δ and δ' are $u_{2n} \ldots u_{4n-1}$ or $u_{2n} \ldots u_{4n}$ and $u_{2n+1} \ldots u_{4n+1}$ or $u_{2n+1} \ldots u_{4n+2}$. In every case they overlap. Then, by induction, it can be shown that $w = (u_1 \ldots u_{n-1})^{-1} w_0$ and then w is not ultimately periodic. Thus C is a weakly prefix code. □

Using the composition proposition 2.5 and the previous examples, it is easy to construct over a finite alphabet examples of codes having the same properties. Let $B = \{a, b\}$ and $\varphi : A \to B^+$ defined by: $\varphi(u_i) = a^i b$. The language $D = \varphi(A)$ is a prefix code.

Example 2.5 *Let* C *be the code defined in example 2.3. the language* $C' = \varphi(C)$ *is a weakly prefix code over a finite alphabet satisfying:*
    – every word has a countable (finite or not) number of C'-*factorizations*
    – infinitely many words have a countable infinity of C'-*factorizations.*

Example 2.6 *Let* C *be the code defined in example 2.4 the language* $C' = \varphi(C)$ *is a weakly prefix code over a finite alphabet and there exists a word:* $\varphi(w_0)$ *which has a noncountable infinity of* C'-*factorizations.*

3 The rational case

When a language C is rational, one can consider an automaton $\Omega_0 = (Q_0, q_0, q_F)$ with a finite set of states Q_0, a single initial state q_0 and a single final state q_F, which recognizes C and such that no edge comes to q_0 and no edge goes from q_F. The automaton Ω_0 can be chosen trim (i.e. for every state q there exist a path from q_0 to q and a path from q to q_F) and unambiquous (i.e. the words of C have a single acceptance path). The automaton $\Omega = (Q, q_0, q_0)$ obtained by identification of q_0 and q_F recognizes C^*. If C is a code, the automaton Ω is unambiguous [1]. This automaton looked as a Büchi automaton recognizes C^ω.

Theorem 3.1 *Let C be a rational language $\subset A^+$. The following conditions are equivalent:*

    1. C is a code

    2. every infinite word has a finite number of C-factorizations

    3. there exists p such that every infinite word has at most p C-factorizations.

Proof. $3 \Rightarrow 2$: clear. $2 \Rightarrow 1$: This comes from theorem 1.7. $1 \Rightarrow 3$: Let C be a rational code and $\Omega = (Q, q_0, q_0)$ an unambiguous automaton for C^* constructed as said before. Consider $w \in C^\omega$ and $t \geq 1$. We call <u>cut of (w, t)</u> every sequence $(n_1, \ldots, n_{p-1})$ such that there exists n_p satisfying:

    (i) $n_0 = 0 < n_1 < \ldots < n_{p-1} \leq t < n_p, p \geq 2$ and $w[n_{i-1}, n_i[\in C$ for $i = 1, \ldots, p$. (Here, and in the sequel, the factor $w_i w_{i+1} \ldots w_{j-1}$ of a word w is denoted by $w[i, j[$).

    At first, we show that, for every t, (w, t) has at most $\text{Card}(Q)$ cuts.

    Let us consider $(n_1, \ldots, n_p)$ and $(n'_1, \ldots, n'_k)$ such that (i) is satisfied. Denote by q (resp. q') the state reached after reading $w[0, t[$ in the single successful path of $w[0, n_p[$ (resp. $w[0, n'_k[$). If $q = q', w[0, n_p[$ has a second successful path: path related to $w[0, n'_k[$ until t, path related to $w[0, n_p[$ after. Then $p = k$ and $(n_1, \ldots, n_{p-1}) = (n'_1, \ldots, n'_{p-1})$ since Ω is unambiguous.

    Thus (w, t) has at most $\text{Card}(Q)$ cuts. Then w has at most $\text{Card}(Q)$ C-factorisations.   □

Remark: An infinite word which has several C-factorisations is not necessarily ultimately periodic: the word: $ab^2 cb^3 (c^2 b^3) cb^3 (c^2 b^3)^2 \ldots cb^3 (c^2 b^3)^n cb^3 (c^2 b^3)^{n+1} \ldots$ has two C-factorisations when $C = \{a, ab, bcb^2, bc^2 b^2, b^2 cb, b^2 c^2 b\}$.

    A set of infinite words over an alphabet A is said to be rational if it is a finite union of sets $R_i S_i^\omega$ where R_i and S_i are rational subsets of A^*. It was proved that the rational sets of infinite words are the languages which can be recognized by a finite Büchi-automaton [4]. The set of rational subsets of A^ω is closed by finite union, finite intersection and complement [4]. For details, one can see [18].

Proposition 3.2 *Let C be a rational language $\subset A^+$. The set of infinite words which have several C-factorizations is rational.*

Proof. If C is rational, the semi-congruence defined by:

$$u \simeq v \Leftrightarrow u^{-1}C = v^{-1}C$$

is of finite index. Let us denote by $[u]$ the class of a word u. The set D of infinite words which have several C-factorizations can be written:

$$D = \bigcup_{[u] \subset C} C^* . [u] . (C^\omega \cap ([u]^{-1}C - \{\varepsilon\})C^\omega).$$

So D is rational.   □

Remarks:

    – The set of infinite words which have several C-factorisations is countable when the code C has three elements [10]. It can be noncountable when the code C has

more that three elements. For example, let $C = \{ab, aba, bab^2, b^2ab^2a\}$. Every word of $aba(b^2ab^2a + bab^2ab)^\omega$ has a noncountable infinity of C-factorisations.

– It can be proved from proposition 3.2 that, if C is a rational language, C is an ω-code if and only if all its finite subsets are ω-codes. This property does not hold for nonrational languages as it can be seen for $C = \{ab\} \cup \{ab^n ab^{n+1} | n > 0\}$.

– From proposition 3.2, we obtain the next statement which is a result of Staiger [20]. This statement agrees with the fact that a rational ω-language is specified by the set of ultimately periodic words contained in it [4].

Corollary 3.3 *Any rational weakly prefix code is an ω-code.*

Since it can be checked whether the rational set of infinite words which have several C-factorisations is empty or contains a periodic word, we have the following corollary.

Corollary 3.4 *One can decide whether a rational language is a π-code (resp. a weakly prefix code, or equivalently an ω-code).*

The membership problem for the studied classes of codes is decidable in the rational case. Indeed the result is well known for codes [1], and has been proved for codes with bounded deciphering delay by Cori [6]. This latter result is also a consequence of the next result of Capocelli, and can be also deduced from proposition 3.7.

Capocelli [5] gave a necessary and sufficient condition for a rational weakly prefix code (or ω-code) C to have a bounded deciphering delay. That is:

$$\exists p \geq 0 \forall u \in A^* uC^p A^* \cap C \neq \emptyset \Rightarrow C^+ u \cap C^+ = \emptyset.$$

We give here another condition which obviously is satisfied when the code is finite. In this condition we need the notion of *adherence* [3]. An infinite word w belongs to $Adh(C)$, the adherence of a language C of finite words, if every left factor of w is a left factor of a word of C.

Lemma 3.5 *Let us consider a language $C \subset A^+$.*

1. *if C is a code having a finite deciphering delay, C is an ω-code and $C^\omega \cap C^* .Adh(C) = \emptyset$.*

2. *if C is a rational ω-code such that $C^\omega \cap C^* .Adh(C) = \emptyset$, then C has a bounded deciphering delay.*

Proof.

1. If C is not an ω-code C cannot be a code having finite deciphering delay (proposition 1.9). Thus, let C be an ω-code for which there is some $w \in C^\omega \cap C^* .Adh(C)$. Without loss of generality, we may assume that $w = u_1 u_2 u_3 \ldots = u'_1 u'_2 \ldots u'_p w'$ where $u_i, u'_i \in C$ for every $i, w' \in Adh(C)$ and $u'_1 \neq u_1$ or $p = 0$. Since $w' \in Adh(C)$, for every $d \geq 1$ there exists $v \in C$ such that $u_1 \ldots u_d \leq u'_1 u'_2 \ldots u'_p v$ where $u'_1 \neq u_1$ or $p = 0$. Thus C has not the deciphering delay $(d-1)$.

2. Let C be a rational language and $\Omega = (Q, q_0, q_0)$ an unambiguous automaton for C^* constructed as said before. Let d be the number of states. Assume that C has not he delay d. There exist $n \geq 0, u_0, \ldots, u_d, u'_0, \ldots, u'_n \in C, z \in A^*$ such that $u_0 \ldots u_d z = u'_0 \ldots u'_n$ and $u'_0 \neq u_0$.

There exists a path of label $u'_0 \ldots u'_n$ from q_0 to q_0. Within this path, we denote by q_j the state reached after reading $u_0 u_1 \ldots u_j$. There exist j and $j' > j$ such that $q_j = q'_j$ (we denote $q = q_j$). Then we denote: $y = u_0 \ldots u_j = u'_0 \ldots u'_{m-1} x'$ with $x' < u'_m, x = u_{j+1} \ldots u'_j, x' x x^n = u'_m \ldots u'_{m+h}$, with x^n suffix of $u'_{m+h}, u'_{j+1} \ldots u_d z = x^n u'_{m+h+1} \ldots u'_n$. If $h = 0$, for every $n, x' x^n x^n \in C$ and then $y x^\omega \in C^* . \mathrm{Adh}(C) \cap C^\omega$. If $h \geq 1, y x^\omega$ has two distinct C-factorisations: $u_0, \ldots, u_j (u_{j+1}, \ldots, u_{j'})^\omega$ and $u'_0, \ldots, u'_{m-1}, (u'_m, \ldots, u'_{m+h-1}, v)^\omega$ where $v = (u'_{m+h} x'^{-1}) . (x'^{-1} u'_m)$, thus C is not an ω-code.   □

Lemma 3.5 can be used to derive a new proof of a result in [20]. To this end, we consider A^ω as a topological space defined by the set of open subsets: $E \subset A^\omega$ is open iff $E = W A^\omega$ for some $W \subset A^*$. The closed subsets (i.e. the complements of open subsets) are the languages of the form $\mathrm{Adh}(W)$ for some $W \subset A^*$ [21]. We need here the next classes of the Borel hierarchy. A F_σ-set is a countable union of closed subsets and a G_δ-set is a countable intersection of open subsets.

Corollary 3.6 *When C is a code with a finite deciphering delay, the language C^ω is a G_δ-set.*

Proof. Since $\mathrm{Adh}(C^*) = C^\omega \cup C^* . \mathrm{Adh}(C)$ [13], when $C^\omega \cap C^* . \mathrm{Adh}(C) = \emptyset$ the set C^ω is the difference of the closed set: $\mathrm{Adh}(C^*)$ and the F_σ-set: $C^* . \mathrm{Adh}(C)$, hence C^ω is a G_δ-set.   □

Remark: The tempting assumption $"C^\omega = \cap C^i A^\omega"$ is true for the codes C having a bounded deciphering delay [20] but no longer true for the codes C having a finite (but not bounded) deciphering delay (cf. example 3 of [20]).

We can summarise:

Theorem 3.7 *Let C be a rational language $\subset A^+$. The following conditions are equivalent:*
  – C is a code with a bounded deciphering delay
  – C is a code with a finite deciphering delay
  – C is an ω-code satisfying $C^\omega \cap C^* . \mathrm{Adh}(C) = \emptyset$.
  – C is a weakly prefix code satisfying $C^\omega \cap C^* . \mathrm{Adh}(C) = \emptyset$.

We have already seen that there exist ω-codes without finite deciphering delay. The other condition: $" C^* . \mathrm{Adh}(C) \cap C^\omega = \emptyset"$ is neither sufficient. For example, the finite code $\{a, ab, bb\}$ is not an ω-code. Unfortunately proposition 3.7 is false when C is not rational. For example, let $C = \{ab^n c^n d | n \geq 0\} \cup \{a\} \cup b^* c$. Since $\mathrm{Adh}(C) = b^\omega \cup ab^\omega$, the ω-code C satisfies $C^\omega \cap C^* . \mathrm{Adh}(C) = \emptyset$, but the word a has no finite deciphering delay.

In the aim to be complete, let us now observe the finite case. The finite case is almost similar to the rational case. However proposition 3.7, as the result of Levenshtejn [12] and Capocelli [5], show that, in the finite case, the notion of ω-code and the notion of code with bounded deciphering delay coincide. This fact is also a result of Blanchard [2] which uses another notion of factorization ("découpage").

Nevertheless there are a lot of modifications when one considers two-element codes. Indeed, if $\{u, v\}$ is a code, $\{u, v\}$ is also an ω-code [10]. Since the examples given in this paper are chosen with three elements when it is possible, the obtained or recalled results can be recapitulated in the following proposition where A_R (resp. A_F, A_2, A_3) denotes the class of rational (resp. finite, two-element, three-element) languages belonging to a given class of languages A.

Proposition 3.8 *One has the following strict inclusions and equalities:*

$$\mathbf{B} \subset \mathbf{D} \subset \mathbf{I} \subset \mathbf{W} \subset \Pi \subset \mathbf{C}$$

moreover $\mathbf{B_R} = \mathbf{D_R}$ *and* $\mathbf{I_R} = \mathbf{W_R}$ *for rational sets,* $\mathbf{B_F} = \mathbf{D_F} = \mathbf{I_F} = \mathbf{W_F}$ *for finite sets and* $\mathbf{B_3} = \mathbf{D_3} = \mathbf{I_3} = \mathbf{W_3}$ *for three element sets, and finally* $\mathbf{B_2} = \mathbf{D_2} = \mathbf{I_2} = \mathbf{W_2} = \Pi_2 = \mathbf{C_2}$ *for two element sets.*

References

[1] J. Berstel and D. Perrin, *Theory of codes* (Academic Press, Orlando, 1985).

[2] F. Blanchard, Codes engendrant certains systèmes sofiques, *Theoret. Comput. Sci.* **68** (1989) 253-265.

[3] L. Boasson and M. Nivat, Adherences of languages, *J. Comput. System. Sci.,* **20**, (1980), 285-309.

[4] J.R. Buchi, On a decision method in restricted second-order arithmetic, *Proc. Congr. Logic,* Stanford Univ. Press, Stanford (1962) 1-11.

[5] R.M. Capocelli, Finite decipherability of weakly prefix codes, in *Algebra, Combinatorics and Logic in Computer Science* Coll. Math. Soc. J. Bolyai 42, North Holland, Amsterdam (1985) 175-184.

[6] R. Cori, Codes à délai borné maximaux, in *Théorie des codes* Publications du LITP, Universités de Paris VI et Paris VII (1979) 57-74.

[7] J. Devolder, Precircular codes and periodic biinfinite words, *Information and Computation,* **107** n° 2 (1993) 185-201.

[8] J. Devolder, Comportement des codes vis-à-vis des mots infinis et bi-infinis, in *Théorie des Automates et Applications,* (Edit. D. Krob, Rouen, 1991) 75-90.

[9] J. Devolder and E. Timmerman, Finitary codes for biinfinite words, *RAIRO Info. Theor. Appl.,* **26** n° 4 (1992) 363-386.

[10] J. Karhumäki, On three-element codes, *Theoret. Comput. Sci.* **40** (1985) 3-11.

[11] J.L. Lasses, Circular codes and synchronisation, *Internat. Journal of Computer and syst. Sci.* **5** (1976) 201-208.

[12] V.I. Levenshtejn, Some properties of coding and self adjusting automata for decoding messages, *Problemy Kibernetiki,* **11** (1964) 63-121.

[13] R. Lindner and L. Staiger, Algebraische Codierungstheorie, *Theorie der sequentiellen Codierungen,* Akademie-Verlag, Berlin (1977).

[14] I. Litovsky, Générateurs des langages rationnels de mots infinis, Thèse Univ. Lille I (1988).

[15] I. Litovsky and E. Timmerman, On generators of rational ω-power languages, *Theoret. Comput. Sci.* **53** (1987), 187-200.

[16] A. de Luca and A. Restivo, On some properties of very pure codes, *Theoret. Comput. Sci.,* (1980), 157-170.

[17] R. McNaughton, Testing and generating infinite sequences by a finite automaton, *Information and control,* **9** (1966), 521-530.

[18] D. Perrin et J.E. Pin, Mots infinis, *Publications du LITP,* Univ. Paris VI et VII, **91.06** (1991).

[19] M.P. Schutzenberger, Une théorie algébrique du codage, Séminaire Dubreil-Pisot **15** (1955-56), Institut Henri Poincaré, Paris.

[20] L. Staiger, On infinitary finite length codes, *RAIRO Theor. Inform. and Applic.* **20** (1986) n° 4, 483-494.

[21] L. Staiger and K. Wagner, Automatentheoretische undautomatenfreie Charakterisierungen topologischer Klassen regulärer Folgenmengen, *Elektron. Inform.-Verarb. u. Kybernetik, EIK* **10** (1974), 379-392.